An Interfaces Approach to TES Ground Data System Processing Design with the Science Investigator-led Processing System (SIPS)

Rose-Marie Kurian
rose-marie@sdsio.jpl.nasa.gov

Andy Griffin
griffin@canopy.stx.com

Raytheon ITTS/Jet Propulsion Laboratory
299 N. Euclid Avenue, Suite #500
Pasadena, California 91101

Abstract— During the five-year mission, the Tropospheric Emission Spectrometer is expected to produce approximately 8.4 TB of raw data and an additional 25 TB of processed data each year. Developing production-quality software to process the large volumes of scientific data is the responsibility of the TES Ground Data System, which is being developed at the Jet Propulsion Laboratory together with support contractor Raytheon/ITSS. The large data volume and processing requirements of TES pose significant challenges to the design. Selection of the Science Investigator-led Processing System (SIPS) over baselined EOS Data and Information System (EOSDIS) Core System (ECS) will provide the TES project with a cost effective, flexible system, capable of effectively addressing TES data processing challenges. Critical to the design, and the success of processing TES large amount of data successfully, are well defined, well understood, and well managed interfaces.

TABLE OF CONTENTS

1. TES MISSION OVERVIEW

The TES experiment consists of both the instrument and the ground data processing system. The primary scientific objectives of TES are to investigate the 3-dimensional distribution of gases important to tropospheric chemistry, troposphere-biosphere interactions, and troposphere-stratosphere exchange. These investigations will be carried out on global, regional and local scales. The resultant data will be used to calibrate models of the present and future state of the Earth's lower atmosphere. [1]

The TES instrument has two basic science operating modes: Global Surveys (GS) and Special Research Observations. Both the GS and the Special Observation modes contain nadir and limb views.
Both nadir and limb viewing modes are expected to generate massive amounts of data during the mission. This TES data will be collected and processed to produce standard data products that will be distributed to the scientific community on a regular basis. Processing this data is the responsibility of the TES Ground Data System (GDS).

2. TES GROUND DATA SYSTEM OVERVIEW

The TES Ground Data System is part of the EOS Ground System. The interactions between the EOS Ground System and the TES ground system will primarily be via the Distributed Active Archive Center (DAAC) at Langley, VA. The EOS Data and

Operations System (EDOS) is responsible for data capture from the spacecraft and interface of uplink commands. In addition, it will provide the Level 0 data to the DAAC. The Langley Research Center DAAC (LaRC DAAC) will be the focal point for collecting and archiving the externally generated data sets required for TES standard data processing and for archiving and distributing the TES standard data products. *Figure 1* provides a context diagram of the TES GDS.

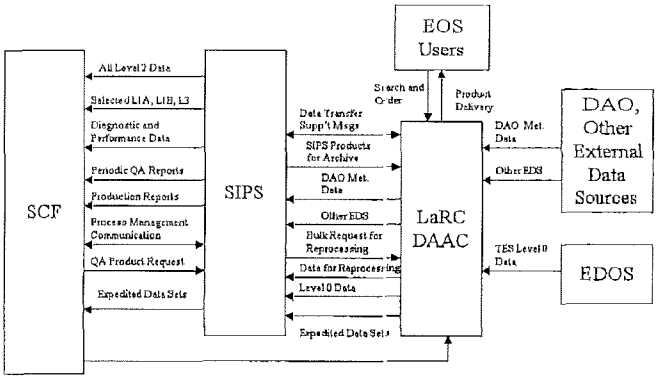

Figure 1 – TES Ground Data System context diagram

The TES Ground Data System (GDS) has three main functional requirements: 1) Routine processing of the Global Survey (GS) Observations, 2) Special Research Observations processing, and 3) Mission Planning. These requirements are allocated as follows. The Scientific Investigator-led Processing System (SIPS) will mainly provide routine GS data processing. Currently, SIPS is being developed at Raytheon ITSS, Pasadena. Special Observations and research activities will mainly be conducted at the Jet Propulsion Laboratory (JPL) Science Computing Facility (SCF). The Science Team is an integral part of the entire TES ground system; they play a key role at the SCF leading the development, data analysis, research and mission planning. Mission Support and planning will also be conducted at the SCF. This will be the responsibility of the Science Team together with the TES Mission Operations (MO) team.

The heart of the TES GDS is the Science Data Processing System (SDPS). The TES SDPS is organized around the TES standard products. The TES SDPS main input is the Level 0 data. The Level 0 data is the standard data set used in both production processing and special processing. The standard science data collection is called a Global Survey and the collection of special science research data is called Special Observation. The TES SDPS accepts both Global Survey and Special Observation data; the processing of these data is divided into subsystems (also called levels) 1 through 3 according to a schema prescribed by the EOS project. The Level 1 processing is broken into Level 1A and Level 1B. Level 1A provides telemetry packet decommutation and a geo-referenced interferograms. Level 1B provides a radiometrically and spectrally calibrated spectrum as a standard data product that is corrected for instrument effects. For each observation, Level 2 performs a "retrieval" of temperature and molecular abundance profiles for a select set of standard species. Level 3 accepts Level 2 Full State Vector files and creates species profiles global maps.

Dividing the TES SDPS into subsystems and further into Product Generation Executables (PGE), defined as a single binary executable, creates a desirable and flexible design that is easy to run and to maintain. *Figure 2* shows the top-level architecture of the SDPS. An important subsystem of the TES SDPS is Framework. The TES Framework is a set of high-level classes that offer functionality common to all PGEs. More specifically, it provides PGE infrastructure, parameter handling, exception handling, File I/O and class hierarchies. [2] [3]

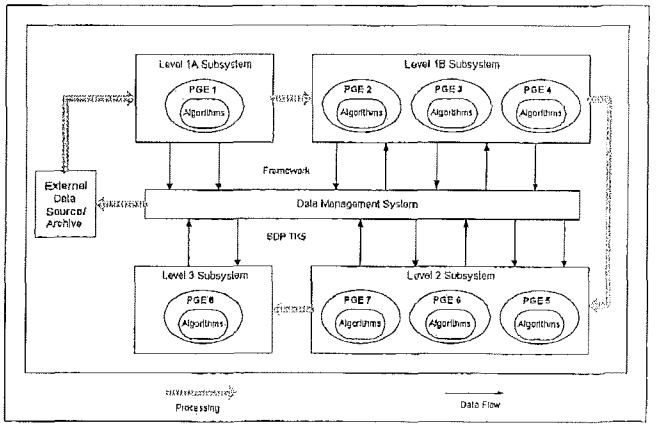

Figure 2 – Science Data Processing System

3. SIPS OVERVIEW AND INTERFACE S

The SIPS is an integrated system that controls the production, short-term archive and distribution of data. It receives inputs from external sources, generates new data via plug-in algorithm packages (in this case PGEs), archives any or all of the information

produced, and manages the distribution of the data. We can think of the SIPS as being analogous to an automated factory that receives, processes, stores and distributes a product. The receiving takes place at the ingest interface while the distribution is accomplished via the subscription interface. The management of the factory, and this case the SIPS, is provided via operations controls through the Operators interface. *Figure 3* provides some detail on the main components of the SIPS and its external interfaces.

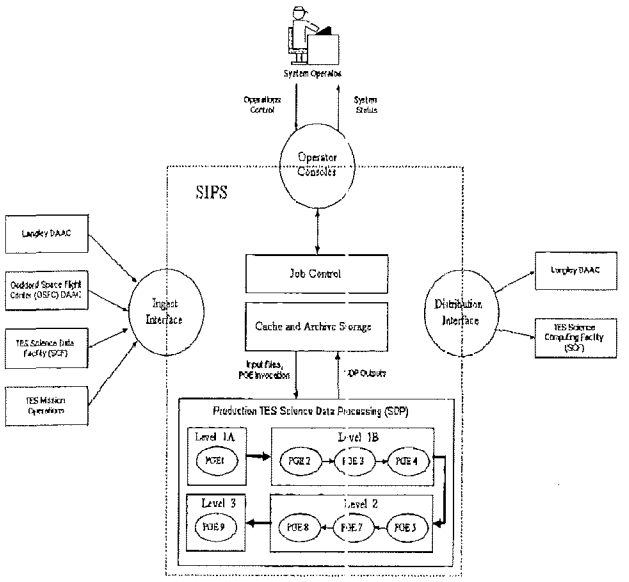

Figure 3 – SIPS Components and External Interfaces

The TES Project is responsible for generating short- and long-term operations plans for the instrument. The Instrument Planning Operations is co-located at the SCF; it consists of a couple of computers, known as the Instrument Support Terminals (IST), with Instrument Support software provided by ESDIS and augmented by mission planning software developed or procured by the TES project. The IST will be used to plan the TES mission and to monitor instrument performance. The Mission Operations (MO) team will have the responsibility of translating the science team mission priorities into instrument sequence uploads.

4. SIPS INGEST INTERFACES

A. LaRC DAAC to SIPS Ingest Interface

The ingest Interface between the LaRC DAAC and the SIPS is automated and well define [4]. The SIPS sends a *Subscription Request* via e-mail for a specific data set. The DAAC will then sent a *Subscription Acknowledgment* to indicate that the subscription has been received. The LaRC DAAC pushes files matching subscription requests to a predefined location at the SIPS (the PDR server). A *distribution notice (DN)* is created supplying the description of the files available. The *distribution notice* is e-mailed to the SIPS. The files are ingested after receiving the *DN*. At the completion of ingest, the SIPS sends a *Production Acceptance Notification (PAN)* detailing success or failure of the file ingest. The following are the main files from the LaRC DAAC to the SIPS.

- *Level 0 Data*: The Level 0 data, also known as the Production Data Set (PDS), or high rate data since it is transmitted through the high rate interface. This data comes to the SIPS from EDOS via the LaRC DAAC where they are also archived. Level 0 data is organized into four types of PDS data: science data, state data, performance data and diagnostic data. All level 0 data is in the Earth Science Data Type (ESDT) format.
- *Expedited Data Sets (EDS)*: Expedited data sets are used to monitor the instrument's on-orbit performance and received minimal processing from EDOS before being transferred to the LaRC DAAC for archival and distribution to the SIPS. The mission operations team may request EDS for up to 2% of the total TES data volume to be automatically delivered to the LaRC DAAC. In addition, the mission operations team may request data from a specific contact. Expedited Data Sets are also of the same four types as the PDS data listed above.
- *Meteorological data*: DAO or ECMWF data that is required for SDPS processing both at the SIPS and SCF.
- *External-Operational Support Products (OSP)*: These are external data sets needed for use in the science data processing and are used without any modifications. Some examples of possible external OSPs are NCEP data, OMI ozone data, TOMS Ozone data and ASTER Surface Model data.
- *Archive data for Reprocessing*

B. Goddard Space Flight Center (GSFC) DAAC to SIPS Ingest Interface

The ingest Interface between the GSFC DAAC and the SIPS also consists of both a subscription and FTP transfer mechanism as described above in the LaRC

DAAC to SIPS interfaces. The following files will arrive to the SIPS from GSFC DAAC and a copy will be sent to the LaRC DAAC for archival.

- *Pre-processed DPREP files:* Aura Spacecraft Definitive Ephemeris and Aura Spacecraft Refined Attitude Data

C. Science Computing Facility (SCF) to SIPS Ingest Interface:

Interface between the SCF and the SIPS can be as follows: (1) Via the *Configuration Management (CM) System* (all software deliveries and Operational Support Products will be delivered through the configuration management system), (2) *Electronic mail* (e-mail will be use to send notifications that SCF generated data products, such as special observations and metadata quality assessment files, are staged at an agreed upon location and are ready to be picked-up using ftp transfers. In addition request to receive special observations data files will also be done via e-mail) and (3) *Meetings* (management of the SIPS processing will be through meetings with representations from the Science team, SIPS team and Mission Operations team). The main files between the SCF and the SIPS are:

- *SDPS Software Deliverables*: this will be done per release and as needed when changes are implemented. These deliveries will be formal through the Configuration Management System and will include a Release Description Document (RDD), PGE Definitions, Executables and Installation Procedures.
- *SCF Generated Products*: send to the SIPS for archiving purpose (FTP to SIPS with e-mail notification)
- *SCF Special Observations Data*: sent to the SIPS for archival at the SIPS or at the LaRC DAAC
- *Special Observation ordering*: the SCF (science team) will have special interest in location specific special observations for further research. These specific data can be ordered from the SIPS via e-mail.
- ECS Metadata Control Files (MCF): one per ESDT (via CM system)
- *Operational Support Products*: these are files and/or databases essential for TES SDPS processing created at the SCF and delivered to the SIPS via the configuration management system.
- *QA Product Request*: the science team at the SCF requests some sample or a selected group of the QA products via e-mail to the SIPS.
- *QA Metadata Update*: any metadata updates necessary after further data analysis are sent via e-mail to both SIPS and DAAC
- *Process Management Communication*: global survey observation processing planning is expected to be an infrequent activity once a baseline is operational. Most of the science planning will be for the Special Research Observations.

D. TES Mission Operations to SIPS interface:

5. SIPS DISTRIBUTION INTERFACES

A. SIPS to LaRC DAAC Distribution Interface

All standard data product will be sent to the LaRC DAAC through automated FTP transfers. The SIPS puts their data on the SIPS PDR server together with a *Product Delivery Record (PDR)* notice. LaRC DAAC polls the PDR server and when it detects a PDR file it "pulls" the PDR notice and all the related files. If there is an error in the transfer, or a discrepancy between the PDR notice and the files, the LaRC DAAC sends a *Product Delivery Record Discrepancy (PDRD)* notice to the SIPS. If, on the other hand, the transfer is successful and the files transferred match the PDR notice then the LaRC DAAC will issue a *Production Acceptance Notification (PAN).*

- *Level 1 Standard Data Products*: Engineering Data, PCS/ICS Data, Spacecraft Position Data, Error Sink, Both Nadir and Limb Spectra (calibrated and phase corrected) Nadir spectra, and cold and black body Spectra Calibration. [5]
- *Level 2 Standard Data Product*: these 14 files are vertical volume mixing ratios of the following selected molecular species: H2O (1 nadir and 1 limb), O3 (1 nadir and 1 limb), CH4 (1 nadir and 1 limb), CO (1 nadir and 1 limb), NO2 (limb), NO (limb), HNO3 (limb), atmospheric temperatures (1 nadir and 1 limb) and Ancillary data. [5]
- *Level 3 Standard Data Products*: global maps of the following species: H2O (1 nadir and 1 limb), O3 (1 nadir and 1 limb), CH4 (1 nadir and 1 limb), CO (1 nadir and 1 limb), NO2 (limb), NO (limb), HNO3 (limb), and atmospheric temperatures (1 nadir and 1 limb) all on selected pressure or isoentropi surfaces to facilitate browsing and

comparison to models and other space-borne observations. [5]

- *Other SIPS Products*:
  1. *Algorithm Package*: this is a tar file of the algorithm code
  2. *QA Updates*: after further analyzing the data the science team at the SCF may have metadata QA updates. These will be sent to the DAAC via the SIPS using the subscription mechanism.
- *Special Observation Data*: this is data that has been processed at the SCF and sent from the SCF to the SIPS for archival at LaRC DAAC
- *Standard Data Product Metadata Files* (*.met files): files containing metadata information (by subscription).

B. SIPS to SCF Distribution Interface

The SIPS and the SCF interface via fiber optic networks provided by the Jet Propulsion Laboratory (JPL). This interface supports all software deliveries, standard data products, QA data, production planning and scheduling, operational support products, and documents between SIPS and SCF. The subscription mechanism between the SIPS and the SCF is not as formal as the one between the LaRC DAAC and the SIPS. In this case, a subscription is initiated by and e-mail sent from the SCF to the SIPS. Once the requested data set is generated SIPS responds back to the subscription by sending an e-mail to the requester indicating that the data is available and where it can be found. Through a web interface the requester can then proceed to retrieve the data. FTP data transfers will occur via a web interface and would be initiated by the requester.

Electronic e-mail will be the preferred method to request any information that is not a subscription. The following files are the main deliverables from the SIPS to the SCF.

- *Level 0 Expedited Data Set (EDS)*: this data is obtained by the SIPS from LaRC DAAC and provided to the SCF through subscription. The SIPS archives this data without processing.
- *Special Observation Data*: special observation data may require special research processing. This will take place at the SCF.
- *Mission Operations Support Data*: such as State data, PCS and ICS Performance data, and Diagnostic data such as Memory and sequence dumps.
- *All Level 2 Data*: this data is use to perform data QA analysis.
- *Selected Level 0, Level 1A, Level 1B, and Level 3*:
- *Diagnostic and Performance Data*:
- *Periodic QA Reports*: are sent in response to the SCF QA processing request.
- *Failed PGEs*: tar files (manual FTP with E-mail notification). This is sent to the SCF for analysis. This tar file will contain the inputs and outputs used in the processing of the failed PGE; export of database tables specific to the failed PGE and production history related the particular PGE run.
- *Production Summary and run logs*: (FTP or e-mail) these are production reports and run logs providing information on the production
- *Production History*: information specific to each PGE run. (FTP or e-mail)
- *Meteorological data*: DAO or ECMWF data that is required for SDPS processing both at the SIPS and SCF (is this an external OSP or OSI???)
- *External-Operational Support Products (OSP)*: These are external data sets needed for use in the science data processing and use without any modifications by the SDPS. Some examples of possible external OSPs are NCEP data, OMI ozone data, TOMS Ozone data and ASTER Surface Model data.
- *Process Management Communication*: global survey observation processing planning is expected to be an infrequent activity once a baseline is operational. Most of the science planning will be for the Special Research Observations.

6. SIPS OPERATOR CONSOLE INTERFACES

The main interfaces through the SIPS Operator console are:

- *Operations Control*: planning and re-planning of global surveys will be an infrequent activity after the initial baseline is establish. Most of the planning will be for Special Observations with monthly updates.
  1. Production plans (from the science team) e-mail
  2. Weekly operations plans: submitted daily via (ftp or e-mail)
- *System Status*: operators can monitor the SIPS through a web interface. They can view the status of all PGEs running and all busy and idle CPUs.

7. OTHER TES GROUND DATA SYSTEM INTERFACES

A. Mission Operations to SIPS and SCF

- *Weekly operations plans*: submitted daily via (ftp or e-mail)

B. ECS (Landover Maryland) to SCF

- *ECS Metadata Control Files (MCF)* – one per ESDT file (via e-mail).

C. EOS Operations Center (EOC) – SCF Interfaces

- *TES telemetry anomaly notification*: notification to the TES Principal Investigator (PI) that a TES telemetry anomaly has occurred.
- *Low rate engineering data*: will be available to the PI via the Instrument Support Toolkit (IST)

D. EOS USERS – LaRC DAAC Interfaces

- *Search and Order*: through the EOS Data Gateway (EDG).
- *Product Delivery*: an e-mail message, with the subject "ECS Notification", will contain the details of how your order will be delivered. Each data file will be delivered with a metadata file that will be identified by the data file name with a *.met* appended.

E. SCF – EOC Interfaces

- Prearranged response sequence to be used in the event of a TES telemetry anomaly.

8. CONCLUSION

Early identification and planning of interfaces can be a key factor in the success of the TES Ground Data System using the SIPS approach to data processing. Processing the data at the SIPS gives the TES project more control over the software and hardware architecture, allowing for a flexible design that can better take advantage of new technologies but with fewer constraints on the science software. This will allow the TES GDS to be able to more effectively process all the collected data.

ACKNOWLEDGEMENTS

The author would like to acknowledge Mark Echeverri Raytheon ITSS systems engineer for the Microwave Limb Sounder (also aboard AURA) SIPS for his valuable inputs. Also, special thanks to Scott Lewicki, systems engineer at the Jet Propulsion Laboratory, for sharing his expertise and knowledge.

REFERENCES

[1] Reinhard Beer, *TES Ground System Operations Concept*, D-18451, Caltech/JPL, CA, 2000.

[2] Steve Larson, Stephen Watson and Kalyani Rengarajan, "A Framework-Based Approach to Science Software Development", *IEEE Aerospace 2000 Conference Proceedings*, March 18-25, 2000.

[3] Stephen H. Watson, Akbar A. Thobhani, Benny B. Chan, "Design and Implementation of the Tropospheric Emission Spectrometer Science Data Processing System Framework", IEEE Aerospace 2001 Conference Proceeding,

[4] Interface Control Document between the EOSDIS Cor System (ECS) and the Science Investigator-le Processing System (SIPS), Volume 0, 423-41-75-00

[5] Interface Control Document between the EOSDIS Core System (ECS) and the Science Investigator-led Processing System (SIPS), Volume 10 – Tropospheric Emission Spectrometer (TES), 423-41-57-10

Rose-Marie Kurian is a software system engineer for the TES Ground Station System at Raytheon ITTS, Pasadena. She is also the Technical lead for the TES team members working at Raytheon. She has extensive experience in remote sensing programs. Rose-Marie has a M.S. degree in applied mathematics, and a M.S. degree in managing information system.

Andy Griffin is a software system engineer of Raytheon ITSS. He has extensive experience in archive and distribution system